Occupational Hygiene, Vol. 3, pp. 169–176
Reprints available directly from the publisher
Photocopying permitted by license only

RECOMMENDATIONS FOR REDUCING THE EFFECTS OF EXPOSURE MISCLASSIFICATION ON RELATIVE RISK ESTIMATES

MUSTAFA DOSEMECI and PATRICIA A. STEWART

Occupational Studies Section, National Cancer Institute, Rockville, MD, USA

Misclassification of exposure can severely affect estimates of disease risks and even, in some extreme situations, result in misleading interpretations of exposure-disease associations. Although the effects of misclassification on risk estimates have been evaluated in several studies, the effect in cohort-type distributions with multiple exposure categories has not been addressed, and no recommendation has been made to exposure assessors for reducing these adverse effects of misclassification. Hypothetical distributions typically observed in occupational cohort and case-control studies were generated in order to evaluate the effects of nondifferential exposure misclassification on risk estimates. We identified four major determinants of exposure misclassification in this hypothetical exercise: (1) the size of the true risk; (2) the amount of misclassification; (3) the prevalence of the true distribution of exposure; and (4) the direction of misclassification, For cohort-type distributions, where the prevalence of exposure is high, extreme distortion was observed when misclassification was from the exposed categories to the unexposed category; little effect was observed when the misclassification was from the unexposed to the exposed categories. For the type of distribution seen in case-control studies, where the exposure prevalence is low, greater effects were observed when misclassification was from the unexposed category to the exposed categories, while little effect was observed when the misclassification was from the unexposed category to the exposed ones or between exposed categories. On the basis of these results and our experience in assessing exposure for epidemiological studies, we have developed recommendations to exposure assessors for reducing the adverse effects of exposure misclassification on relative risk estimates.

INTRODUCTION

Misclassification of exposure is one of the factors that is crucial in distorting an estimate of disease risk and may result in misleading interpretations of exposure-disease associations. Although a number of studies have addressed the effects of misclassification on risk estimates (Marshall *et al.*, 1981; Armstrong and Oakes, 1982; Flegal *et al.*, 1986; Dosemeci *et al.*, 1990; Wacholder *et al.*, 1991; Stewart and Correa-Villasenor, 1991; Birkett, 1992; Brenner, 1992; Brenner *et al.*, 1992; Kauppinen *et al.*, 1992; Brenner and Blettner, 1993; Brenner *et al.*, 1993), few have addressed the effect of misclassification in multiple exposure categories (Marshall *et al.*, 1981; Armstrong and Oakes, 1982; Dosemeci *et al.*, 1990; Birkett, 1992). The objectives of this study are to describe the determinants of nondifferential misclassification of exposure and its ef-

For Correspondence: M. Dosemeci, Occupational Studies Section, National Cancer Institute, Building EPN Room 418 Rockville MD 20892, USA.

fects on risk estimates, and to develop recommendations to occupational hygienists to minimize the adverse effects of misclassification.

METHODS

Hypothetical distributions typically observed in occupational cohort (85% exposure prevalence = 85% exposed, 15% unexposed) and case-control studies (15% exposure prevalence = only 15% exposed) were generated in order to evaluate the effects of non-differential exposure misclassification on risk estimates. Three exposure categories (none, low and high) were used with various patterns of misclassification of exposure to measure the effects of various determinants of exposure misclassification on risk estimates. We identified four major determinants of exposure misclassification in this hypothetical exercise: (1) the size of the true risk, that is, the actual value of the true relative risk (RR) (RR = 2, 4 or 8); (2) the amount of misclassification (10, 30 or 50%); (3) the prevalence of the true distribution of exposure (1–99%); and (4) the direction of misclassification, i.e. misclassification of truly unexposed subjects to the exposed categories, misclassification of truly exposed subjects to the unexposed category, misclassification of subjects between the exposed categories and misclassification between all categories (exposed and unexposed).

In evaluating the effects of each determinant on the risk estimate, a range of the determinant under evaluation was used, the other determinants remaining constant. The constants were: for the size of the true risk, RR = 2.0 and 4.0 for the low and high categories, respectively; for the amount of misclassification 30%; for exposure prevalence, 50%; and for the direction of misclassification, misclassification of subjects among all exposure categories.

Distorted distributions were calculated from the true distributions by applying the appropriate misclassification pattern. The distorted RRs were then calculated and compared with the true RRs.

RESULTS

The effects of the four determinants of misclassification on risk estimates are presented in Figures 1–4. For comparison, the true and the distorted relative risks are presented in the same figure. Figure 1 presents the effect of the size of the true risk on the risk estimates, ranging from 1.5 to 8.0, using an exposure prevalence of 50% and a 30% misclassification in every direction (10% from unexposed to exposed categories; 10% from exposed to unexposed categories; and 10% between exposed categories). Distortion towards the null value increases as the size of the true RR increases, indicating that larger RRs are more sensitive to exposure misclassification than smaller ones.

The effect of the amount of misclassification is presented in Figure 2. As expected, the magnitude of the bias towards the null increases with increasing amount of misclassification, indicating the importance of accuracy in exposure assessment.

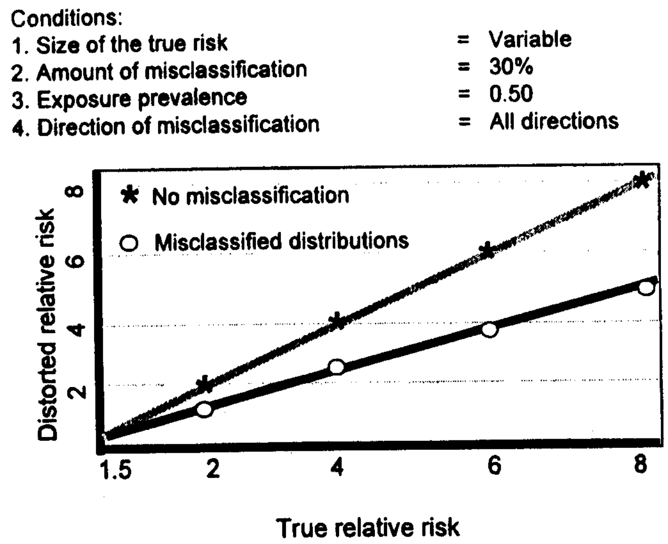

Figure 1 Effects of the size of the true risk on the risk estimate

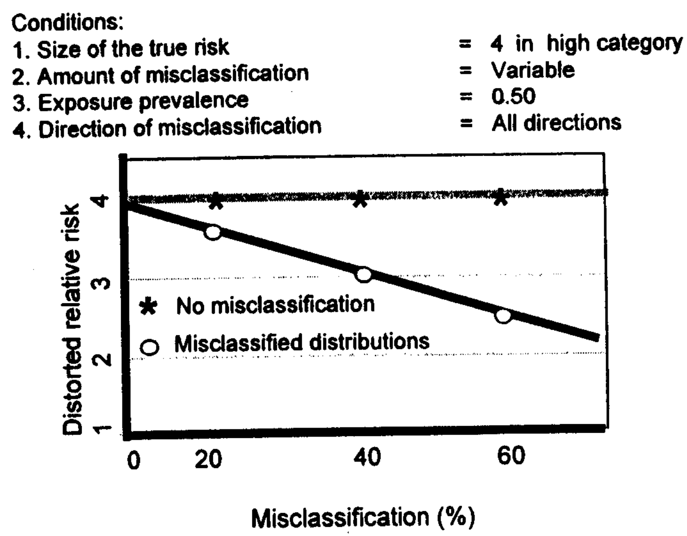

Figure 2 Effects of the amount of misclassification on the risk estimate

An interesting result is the effect of exposure prevalence on the risk estimates (Figure 3). When misclassification occurs in every direction, the minimal amount of distortion of the RR is observed when the exposure prevalence is 50%. The maximal distortion is observed at the two extreme ends of the prevalence, that is, when the exposure prevalence approaches 0 or 100%. Thus, risk estimates that could be generated from both case-control (1–15% exposure prevalence) and cohort (85–100% exposure prevalence)

Conditions:
1. Size of the true risk = 4 at high category
2. Amount of misclassification = 30%
3. Exposure prevalence = Variable
4. Direction of misclassification = All directions

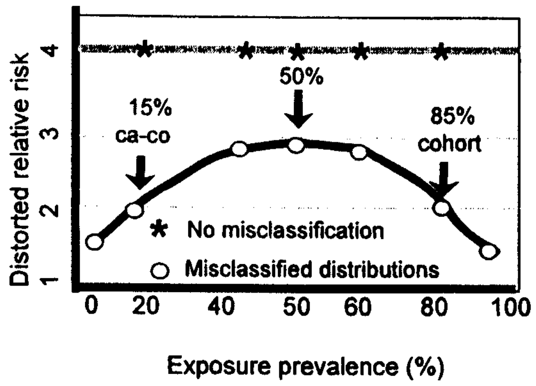

Figure 3 Effects of prevalence of the true distribution of exposure on the risk estimate

study designs are strongly affected by misclassification when the direction of that misclassification is in all directions.

The most enticing finding is the effect of the direction of misclassification on risk estimates. The four most likely patterns of misclassification are presented in Figures 4a–d. Because the effect of the direction of misclassification is closely related to the true distribution of the exposure prevalence, we evaluated the effects for two typical prevalences of exposure, i.e. 15% to represent case-control studies and 85% to represent cohort studies. Figure 4a shows the pattern when misclassification occurs from the unexposed category to exposed categories; e.g. 20% truly unexposed subjects are misclassified in the low-exposure category and 10% of the unexposed subjects to the high-exposure category. Under these conditions, case-control studies (15% exposure prevalence) show a large bias towards the null value, while cohort studies (85% exposure prevalence) show very little effect. In Figure 4b, the direction of misclassification was the opposite, that is 20% of the subjects with low exposure and 10% of those with high exposure were moved to the unexposed category. The RRs for the cohort design are highly biased towards the null value, while the effect in the case-control design was minimal. When misclassification of subjects occurs only between the exposed categories (Figure 4c; i.e. 15% of the subjects were moved from the low-exposure category to the high category and 15% from the high-exposure category to the low one), the risks are overestimated (i.e. away from the null) in the low-exposure category and underestimated (towards the null) in the high-exposure category in both the case-control and cohort designs. Finally, when misclassification occurs in all directions (Figure 4d; i.e. 7% of subjects in the unexposed category are moved to the low-exposure category and 3% to the high-exposure category; 5% of subjects in the low-exposure cat-

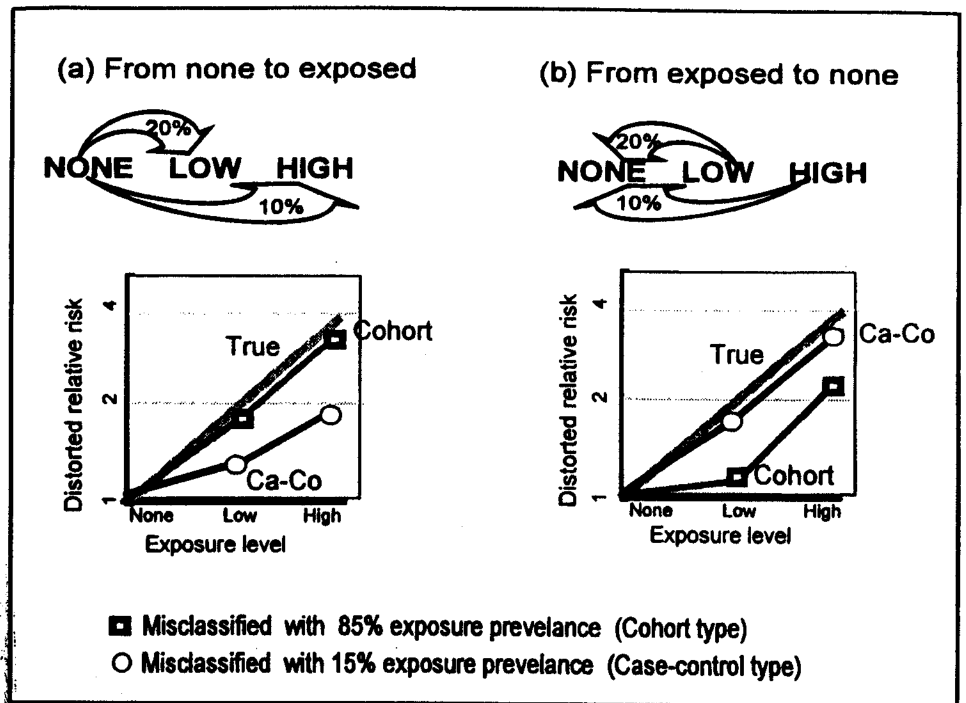

Figure 4 Effects of the direction of misclassification: (a) from unexposed to exposed; (b) from exposed to unexposed; (c) between exposed categories; (d) in all directions

egory are moved to the unexposed and to the high-exposure categories; and 7% of subjects in the high-exposure category are moved to the low-exposure category and 3% to the unexposed category), the risks are underestimated towards the null value in both case-control and cohort designs.

DISCUSSION

In this hypothetical exercise, we evaluated the effects of four determinants of exposure misclassification on risk estimates. As expected, the bias increased with increasing size of the true risk and the amount of misclassification. One interesting finding was the effect of exposure prevalence on risk estimates: a U-shaped curve was seen, in which the bias first decreased, as the exposure prevalence approached 50%, and then increased, as the exposure prevalence increased. A similar U-shaped curve was observed in 2 × 2 table analyses (Flegal *et al.*, 1986). Minimal bias was observed when the exposure prevalence was 50%. Because the effect of exposure prevalence is dependent on the direction of misclassification, we selected a misclassification pattern (i.e. in all directions) in which the effect of the direction would be minimal. This pattern was also selected because it is the most usual real-life situation. The large amount of bias at

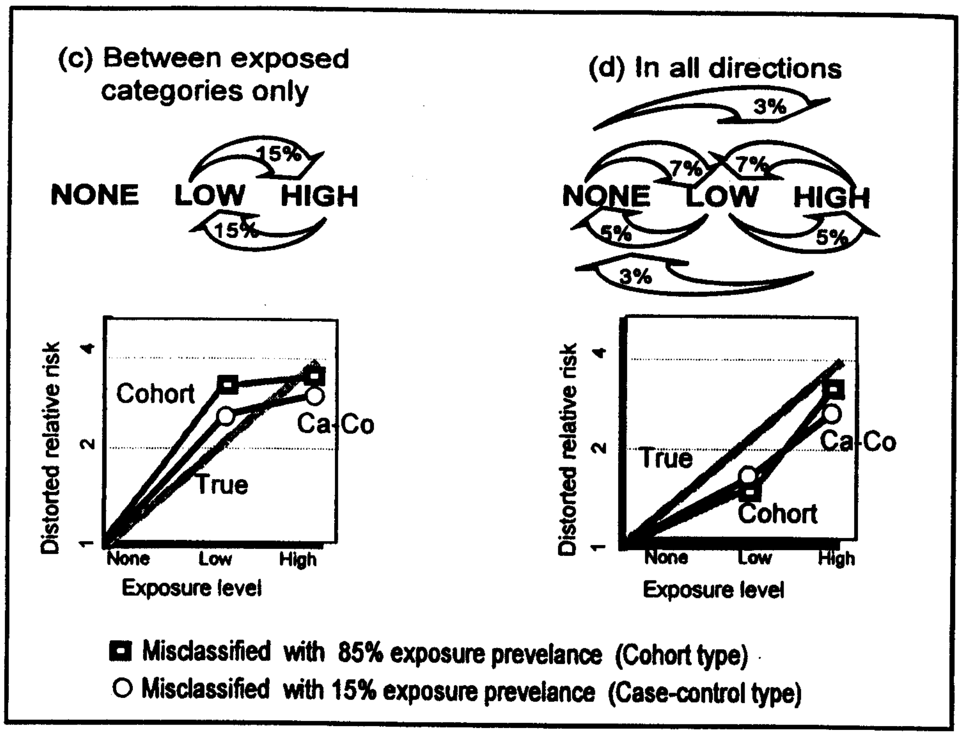

Figure 4 (Continued)

prevalences typically seen in both population-based case-control and cohort studies indicates that there is a strong potential for bias in RRs when misclassification occurs in all directions. One means of reducing this adverse effect of misclassification would be to design a study in which the exposure prevalence is as close as possible to 50%.

Another interesting finding of this exercise is the effect of the direction of misclassification on the different study designs. Case-control studies showed more bias than cohort studies when the misclassification was from the unexposed to the exposed categories. In contrast, this design showed minimal bias when the direction of misclassification was from exposed to unexposed. These findings suggest that if, in the exposure assessment process of a case-control study, where the exposure prevalence is low, an occupational hygienist is not sure about the exposure status of a subject, it is judicious to classify that subject as unexposed. In cohort studies, however, where the exposure prevalence is high, there was more bias than in case-control studies when the misclassification was from the exposed to the unexposed category. In cohort studies, therefore, the strategy should be to classify subjects of uncertain exposure as exposed, in order to minimize the adverse effects of exposure misclassification on risk estimates. This phenomenon has been described (Flegal *et al.*, 1986; Stewart and Correa-Villasenor, 1991) with respect to case-control studies with dichotomous exposure categories, but the effect of misclassification has not been evaluated for polychotomous

Another interesting finding was the effect on risk estimates of misclassification only between exposed categories. In this situation, the observed risk was biased away from the null in the middle exposure category but was biased towards the null in the high-exposure category, resulting in a distortion of the dose-response relationship. This type of risk pattern is often seen in occupational epidemiological studies, and Armstrong and Oakes (1982) predicted similar situations in occupational settings. Such a finding in a study may indicate that the occupational hygienist did a good job in classifying truly unexposed subjects but perhaps, because of limitations of historical exposure information, did a poorer job in assigning the exposed subjects into proper exposure categories. This observation suggests where emphasis should be placed when validating exposure assessments.

Finally, when misclassification occurs in all directions, the RR was biased towards the null value in both case-control and cohort studies, with no important differences between the two study types.

RECOMMENDATIONS

The following recommendations are made on the basis of the results of this hypothetical exercise, in order to reduce the adverse effects of misclassification on risk estimates:

- The effect of misclassification should be considered in interpreting the results of a study in relation to the possible size of the true RR, because the magnitude of misclassification bias depends on the size of the true RR.
- Care should be taken to achieve accurate exposure assessment, because the magnitude of the bias is strongly dependent on the proportion of misclassified subjects.
- In selecting the study population, exposure prevalence should be taken into account relative to the direction of misclassification. If misclassification is expected in every direction, the exposure prevalence should be as close as possible to 50% in order to have a minimal effect of misclassification on the risk estimates. It would be difficult to apply this recommendation in a population-based case-control study but it would perhaps not be very difficult in cohort or nested case-control studies.
- The exposure status of a subject for whom little or no information is available should be classified in relation to the exposure prevalence of the study population. If the subject comes from a population with low exposure prevalence, as in community-based case-control studies, he or she should be classified as unexposed. If the subject comes from a population with a high exposure prevalence, as in cohort studies, he or she should be classified as exposed. In general, if the exposure status of a subject is in doubt, he or she should be placed in the category with the largest number of subjects.
- The exposure assessment must be validated in a subgroup of the study population in order to evaluate the quality of the exposure assignments. The criteria for selcting the subgroup to be validated may be determined by the observed risk pattern. For example, if the risk pattern is a nonlinear dose-response relationship, with a higher risk in the middle exposure category and a lower risk in the high-exposure category, the emphasis of the validation effort should be placed on the exposed group, because the observed dose-response relationship suggests that subjects in the unexposed category have been

classified accurately and that the misclassification probably occurred between exposed categories only.

References

Armstrong, B. G. and Oakes, D. (1982). Effects of approximation in exposure assessments on estimates of exposure response relationships. *Scand. J. Work. Environ. Health*, **8**, 20–23.
Birkett, N. J. (1992). Effect of nondifferential misclassification on estimates of odds ratios with multiple levels of exposure. *Am. J. Epidemiol.*, **136**, 356–362.
Brenner, H. (1992). Notes on the assessment of trend in the presence of nondifferential exposure misclassification. *Epidemiology*, **3**, 420–427.
Brenner, H. and Blettner, M. (1993). Misclassification bias arising from random error in exposure measurement: implications for dual measurement strategies. *Am. J. Epidemiol.*, **138**, 453–461.
Brenner, H., Savitz, D. A., Jockel, K. H. and Greenland, S. (1992). Effects of nondifferential exposure misclassification in ecologic studies. *Am. J. Epidemiol*, **135**, 85–95.
Brenner, H., Savitz, D. A. and Gefeller, O. (1993). The effects of joint misclassification of exposure and disease on epidemiologic measures of association. *J. Clin. Epidemiol.*, **46**, 1195–1202.
Dosemeci, M., Wacholder, S. and Lubin, J. H. (1990). Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am. J. Epidemiol.*, **132**, 746–748.
Flegal, K. M., Brownie, C., and Haas, J. D. (1986). The effects of exposure misclassification on estimates of relative risk. *Am. J. Epidemiol.*, **123**, 736–751.
Kauppinen, T. P., Mutanen, P. O., and Seitsamo, J. T. (1992). Magnitude of misclassification bias when using a job-exposure matrix. *Scand. J. Work Environ. Health*, **18**, 105–112.
Marshall, J. R., Priore, R., Graham, S., and Brasure, J. (1981). On the distortion of risk estimates in multiple exposure level case control studies. *Am. J. Epidemiol*, **113**, 464–473.
Stewart, W. F. and Correa Villasenor, A. (1991). False positive exposure errors and low exposure prevalence in community based case control studies. *Appl. Occup. Environ. Hyg.*, **6**, 534–540.
Wacholder, S., Dosemeci, M., and Lubin, J. H. (1991). Blind assignment of exposure does not always prevent differential misclassification. *Am. J. Epidemiol.*, **134**, 433–437.